More than One Way to Skin a Cat: Why Full-Sentence Definitions Have not Been Universally Adopted

Michael Rundell
Lexicography MasterClass Ltd
252 Wincheap
Canterbury CT1 3TY
U.K.
michael.rundell@lexmasterclass.com

Abstract
At the last Euralex Congress, John Sinclair reiterated the case for full-sentence definitions (FSDs), and questioned why the COBUILD approach to defining had not been generally adopted by other dictionary publishers. This paper answers his question. The theoretical case for FSDs is reviewed (and in general not challenged), and it is shown how the full-sentence model often results in definitions that are more effective and more readable than could be achieved using traditional styles. But the FSD is not always the most appropriate strategy: the approach has several disadvantages, and a rigid adherence to this style does not always serve best interests of dictionary users (especially language learners). Rather, it will be argued, the goals that gave rise to the FSD may often be achieved through other means. The paper concludes with proposals for a range of defining strategies (including FSDs), along with suggestions as to when each is likely to be most effective.

1 Introduction[1]

It is almost 20 years since the first COBUILD dictionary burst upon the scene and changed the face of lexicography. Its most important innovation – the systematic use of corpus data as the foundation of its description of English – has fundamentally altered the way dictionaries are compiled. Despite some initial resistance, corpus lexicography has become the norm, and few lexicographic enterprises in English would now be undertaken without a basis in primary linguistic data.[2] The effects of the corpus revolution are still being felt, as corpora grow larger and more diverse and the software for exploring them becomes increasingly powerful (see e.g. Kilgarriff et al. 2004). Most importantly, corpora are now available (or being developed) for dozens, if not hundreds, of languages, bringing the benefits of data-

[1] Special thanks to my colleague Faye Carney for her insightful comments.
[2] In the U.S., however, resistance to corpus lexicography is still depressingly widespread.

driven lexicography to an ever widening circle. There is no going back – but no-one who has worked with corpus data would ever want to go back.

Among COBUILD's other innovations, the most striking is its use of 'full-sentence definitions' (FSDs). As the name implies, FSDs present defining information in the form of a complete sentence in which the *definiendum* is embedded, for example:

confide If you **confide** in someone, you tell them a secret
confidential Information that is **confidential** is meant to be kept secret or private[3]

While this approach is not entirely original – FSDs can be found in early English dictionaries, from a time when lexicographic conventions had not yet hardened[4] – its use in the COBUILD range is both coherently motivated and systematically applied.

At the last Euralex Congress, John Sinclair (2004) reiterated his faith in this approach, and wondered aloud why FSDs had not been taken up by other dictionaries. It is a good question. A notable feature of lexicography in general – and of English monolingual learners' dictionaries (MLDs) in particular – is the way that any innovation which looks likely to improve the product is sooner or later adopted by competitor dictionaries. Thus, in good Darwinian fashion, the standard model evolves and improves. The use of controlled defining vocabularies or of devices for indicating the frequency of headwords are cases in point, and the most obvious example of this process is the way that corpora were embraced, almost universally, soon after the publication of the first COBUILD dictionary. The fact that FSDs have *not* been universally copied is significant, and the inference must be that dictionary-makers are not entirely convinced of their value. In fact, FSDs *are* used to some degree all the dictionaries that share the same market slot as COBUILD, and examples of these will be shown later. But it is true they are not used as the default style, so Sinclair's question deserves an answer.

The rest of this paper will:

- briefly outline the case for FSDs presented by Sinclair (1987, 2004, 2005) and Hanks (1987)
- illustrate the benefits of this approach (comparing COBUILD's FSDs with conventional definitions used elsewhere)
- discuss the drawbacks of FSDs in order to explain why they have not been universally adopted
- conclude with some thoughts about the role and value of FSDs in a larger repertoire of defining strategies

[3] References to COBUILD are to the third (2001) edition unless otherwise specified.

[4] For example:
TRANSMUTATION *of Metals* [with *Alchymists*] or the *Grand Operation* (as they call it) is the Finding of the Philosopher's Stone
TRANSCENDENTAL *Curves* [in the *higher Geometry*] are such as cannot be defined by Algebraical Equations, or...
(both from Bailey's *Dictionarium Britannicum*, 1730)

2 The case for FSDs

Ask any fluent speaker of English to tell you what **temerity** means, and nine times out of ten they will come back with a formulation that includes a phrase like 'if someone has the temerity to do something'. Corpus data confirms that the word has a strong preference for appearing in the pattern '**have the temerity+**TO-infinitive',[5] and in essence this provides the motivation for a defining approach based on complete sentences – an approach that 'places the word being explained in a typical structure' (Hanks 1987:117). Traditional lexicographic practice, with its misconceived insistence on substitutability, simply defines the headword as (for example) 'foolish boldness; rashness' (LDOCE1), leaving the user to slot this explanation into an appropriate context. COBUILD rejects this approach because it fails to take account of an essential (one might say, a 'defining') fact about the word **temerity**.

The case for FSDs entails:

• a rejection of existing lexicographic conventions, since – as Hanks diplomatically argues – 'it may well be that their full significance is lost on many readers' (1987:116);

• a move towards a less technical defining language, which 'is designed to read like ordinary English' (Sinclair 1987.xvi), and which to some extent reflects the folk-defining techniques that a teacher or parent might use to explain an unfamiliar lexical item (cf. Stock 1988);

• above all, a philosophical position regarding the language system (which Sinclair has been elaborating since the 1960s), in which a word's typical environment and behaviour are critical to any account of its semantics. Sinclair rejects the notion (implicit in the design of almost all dictionaries) of words as self-sufficient bearers of meaning, and insists that, for any lexical item, 'the characteristic cotext is part of the meaning, and so is relevant to the definition of the item' (2004:5). Put simply, words almost never occur in isolation, so it makes no sense to define them in isolation.

Thus the adoption of FSDs in COBUILD is supported by clear theoretical and pedagogical arguments. It is worth saying at the outset that this paper does not seek to challenge the broad thrust of these arguments: they seem to me theoretically compelling, and every contact one has with language data shows (increasingly) their validity and relevance. My intention rather is to question whether this model of language *necessarily* implies the superiority of FSDs in all cases.

3 Good things about FSDs

The various defining styles available in the FSD model, and the advantages these offer, are introduced and fully explained in Hanks 1987. There is no need to go over this, but it is worth mentioning a few cases where, it seems to me, FSDs clearly work better than any of the alternatives.

[5] For example, in the diverse 250-million-word corpus I am currently using, this pattern occurs in almost 75% of all instances of **temerity** (only 31 of 118 instances are *not* like this).

I have mentioned before the case of the phrasal verb **lay up** (Rundell 1998:333), which nicely illustrates the advantages of the FSD over conventional styles. Compare:

lay sb up cause sb to stay in bed, not be able to work, etc. (OALD4)
lay up If someone is **laid up** with an illness, the illness makes it necessary for them to stay in bed (COBUILD3)

The COBUILD version is superior in every way: it reads like 'normal' prose (unlike the unnatural formula 'to cause someone to...'); it shows that (in this meaning) **lay up** is almost invariably passive; it specifies the reason *why* people are 'laid up' (illness); and it indicates that the verb is typically followed by a PP with **with** (*laid up with a bad cold* for example). The conventional definition fails on every one of these counts.

Another appealing feature of this style is that it allows for additional information to be added in a second sentence. Here, for example, the add-on enables the dictionary to account for the reciprocal use of the verb:

argue If one person **argues** with another, they speak angrily to each other about something that they disagree about. You can also say that two people **argue**.

COBUILD's definition for the verb **slim** indicates its preference for the progressive form:

slim If you are **slimming**, you are trying to make yourself thinner and lighter by eating less food

Though corpus data shows that **slim** is also common in the infinitive (*trying to slim, in order to slim, helped me to slim*, etc), it is certainly rare in any of the 'usual' tenses. Finally (though this is not intended to be an exhaustive list of the virtues of FSDs), it is worth mentioning definitions like the following, which conveys an idea of the illocutionary force of this expression:

When people refer to **the good old days**, they are referring to a time in the past when they think that life was better than it is now.

What is impressive about all these definitions is that they provide a much fuller picture of the target lexical items, yet without making unreasonable demands on users or requiring them to know any special conventions.

4 Why doesn't everyone use FSDs?

While acknowledging the many positive aspects of FSDs, one needs also to be aware of potential disadvantages. These are, in summary:

- *length*: FSDs are usually longer than conventional definitions, and this has a number of consequences
- *overspecification*: it is a feature of FSDs that they specify typical contextual and colligational preferences, but this can sometimes make definitions unhelpfully restrictive
- *new conventions for old*: ironically, successful interpretation of some types of FSD requires familiarity with a new set of lexicographic conventions

4.1 Length

FSDs are almost always longer than conventional equivalents, often twice as long. Sinclair recognises this, and has a number of counter-arguments (2004.5-6), but these are not altogether convincing. Definition length is a genuine problem because it tends to correlate with:

4.1.1 Reduced coverage

COBUILD dictionaries include significantly fewer headwords than other books of the same type. Whatever publishers may claim, the main advanced learners' dictionaries contain a roughly similar number of headwords – but COBUILD's coverage is far narrower. A like-for-like comparison of five randomly selected stretches of text, amounting to about 3% of the alphabet, shows COBUILD falling well behind its competitors. The aggregate figures are:

MED	OALD7	LDOCE4	COBUILD3
1415	1448	1364	1081

Figure 1. Headword-list comparison of main MLDs

The mean of the other three dictionaries is 1409 headwords; COBUILD's headword count is 328 behind, so it has about 23% fewer headwords than its competitors. These are fairly small samples, and one could speculate that any of the three other dictionaries might end up with the largest headword list on a full count. But they are all the same ballpark, while COBUILD falls well short in each of the five samples. This is unlikely to be an aberration.[6]

These are not trivial differences. It is a perfectly reasonable publishing strategy to say: 'we have fewer headwords, but we tell you a lot more about the words we that *do* include',[7] but if the primary function of A-Z dictionaries is for decoding unfamiliar words, a marked reduction in coverage is bound to lead to fewer successful look-ups.

4.1.2 Increased complexity

'If the style of a dictionary is too difficult or too condensed for users, the work is useless' (Sinclair 1987.xvi). Amen to that: the dense, formulaic language of many earlier dictionaries has few supporters in the MLD community. But being too condensed isn't the only way of being difficult. Problems can also arise when definitions are not condensed enough. Consider for example this definition of the noun **retreat**:

---

[6] Counts were done on these stretches of text: checkbook-chimneypot, heaped-hedgehog, pressure cooker-prime mover, turquoise-tzarism, and the whole of the letter N. Adjustments were made for the fact that OALD and COBUILD have a single-entry structure (where LDOCE and MED divide headwords by POS), so that in the former dictionaries **nobody, preview, prey** count as two headwords, **heavy** and **prime** as three.

[7] This is a design feature of the *Longman Language Activator*, for example. As far as I know, though, COBUILD's publishers have never made such a claim.

A **retreat** is a change in your position when you have decided that you do not want to do what you have agreed or promised to do, usually because it has become too difficult, too expensive, or too embarrassing (COBUILD1)

All of this is true, and there are no difficult words in the definition – but this is a challenging sentence for any learner to process. At 39 words, it is off the scale in terms of standard measures of readability like the Flesch test. Readability tests are no doubt a simplistic metric, but it is undeniable that longer definitions mean a heavier reading load (for users whose linguistic resources are limited), and generally entail increased complexity. Thus the abandonment of traditional conciseness can bring new problems for users, who may go from the frying pan of unpacking a dense, formulaic definition to the fire of processing something two or three times longer.

4.1.3 Problems with anaphora resolution

The FSD style (especially the 'if-definition') has a tendency to create sentences in which the word to which a pronoun refers is not altogether clear, for example:

necessitate If something **necessitates** an event, action, or situation, it makes it necessary

bind If one chemical or particle **is bound** to another, it becomes attached to it or reacts with it to form a single particle or substance

This can cause serious processing problems for learners, especially those whose first language (Japanese, for example) does not use pronouns in the same way as most European languages do.[8]

4.1.4 Prolixity and redundancy

A blanket commitment to FSDs – even where the application of a word is very broad and co-text cannot usefully be specified – sometimes leads to explanations that use a lot of words to say not very much. Take the following definition of **gaseous**:

You use **gaseous** to describe something that is in the form of a gas, rather than a solid or liquid

The equivalent definition in MED is almost identical, but it starts at the word 'in', so is half the length. This is not just an issue of bandying word counts. Language-learners who opt to use a monolingual dictionary commit themselves to processing text in another language: this is a challenging task at the best of times, so we owe it to our users not to force them to read more than they need to. To give another example: COBUILD's marketing literature draws attention to this entry for **fortunate** (presumably as a definition that the publishers wish to recommend):

[8] I am grateful to Akihiko Kawahara, who contributed to MED in its early stages, and drew our attention to this as a potential problem with FSDs.

If you say that someone or something is **fortunate**, you mean that they are lucky

What useful information is the learner being here? Arguably, a single word – and 'lucky' is not an adequate definition of **fortunate**, as we shall see later. Of course, the 'displacement' convention 'If you say...' (on which, more later) is intended to convey more about the way the word is typically used. But whether this is apparent to the average user is a very different matter (see 4.3 below).

A final example, the COBUILD definition for the phrasal verb **bump off**:

To **bump** someone **off** means to kill them

Here again, the definition contains a single content word. Though less long-winded than **fortunate**, it is a weak definition. The two things a learner needs to know about this word are, first, its register (noted by all the dictionaries) and second, that it denotes *murder* rather than just killing (noted in most of the dictionaries): you can't be bumped off by avian flu or in a car crash. To the basic 'murder' definition, LLA usefully sets the word in its most typical context (the murky criminal underworld) by adding:

especially because they [the victim] know about things you have done wrong or are dangerous to you.[9]

All of this casts doubt on the claim that the added length of FSDs is offset by their being more informative: 'Full sentence definitions are longer, to be sure, than their abbreviated counterparts. But they contain more information and present it in an immediately accessible form, rather than in a code which has to be learned anew for each dictionary' (Sinclair 2005.428).[10] The definition of **fortunate** discussed above is much less informative than either of the following:

lucky compared with other people, so that life is always easier or more pleasant for you than for them (LLA)

lucky, especially because you have more advantages than other people (MED)

(Both these dictionaries, unlike COBUILD, divide the word into two senses: about people, and about events and situations). Teasing out differences between close synonyms is never easy, but these definitions at least make a creditable attempt. Supported by examples reflecting recurrent patterns in text (such as 'not everyone is as fortunate as you', 'those less fortunate than ourselves', 'we were fortunate enough to...'), they are, overall, the better entries. Both definitions are conventional in style, but they can hardly be said to be framed in 'a code that has to be learned'.

---

[9] This definition was written about 15 years ago, drawing on a 30-million-word corpus. Having checked it against data an order of magnitude larger, I am gratified to see that it holds up well.

[10] Similarly: 'The structure of the full-sentence definition...provides even more detail than is found in other learner's dictionaries' (Barnbrook 2002.47).

It goes without saying that counter-examples can always be found (there are plenty of excellent definitions in COBUILD), but the evidence for an automatic correlation between longer definitions and greater information value is not persuasive. These are independent variables, and it is not hard to find cases where an FSD is *both* longer *and* less informative than conventional equivalents.

4.2 Overspecification

A common problem with conventional definitions is that they are underspecified – that is, in trying to account for all possible instantiations of a word, they often resort to minimalist formulations that can be slotted into any conceivable context. Compare, for example, these two definitions of the word **absolute** when used as a noun:

Something that is absolute (*American Heritage Dictionary*, 3rd edition,1994)
a value or principle that is regarded as universally valid or which may be viewed without relation to other things (*Oxford Dictionary of English* 2005)

The best that can be said of the first definition is that it is easy to follow and infinitely substitutable. But it generalizes to the point of vacuity: it is, in other words, severely underspecified. The second definition is a 'typification' that reflects the majority of observable uses. As Hanks points out when explaining the COBUILD approach, definitions should be read as 'stating what is normally the case rather than what is necessarily the case' (1987.118), a vitally important distinction. And of course 'stating what is normally the case' only becomes possible when we have access to corpus data.

Problems can arise, however, over the interpretation of 'normally'. This is an issue in definitions of all types, but it is especially acute when the primary objective is to place the *definiendum* in a typical context. The cases of **temerity** and **lay up** (above) are good examples of FSDs that accurately reflect the colligational and collocational preferences of these words. But if some conventional definitions are underspecified, FSDs risk the opposite danger: of *over*specifying typical contexts in ways that may cause confusion.. Some examples follow:

innocence If someone proves their **innocence**, they prove that they are not guilty of a crime

This is COBUILD's sole explanation of the use of **innocence** in the sense of 'not being guilty'. Limiting the context to 'proving' one's innocence gives far too narrow an account of how the word is used. A Word Sketch for **innocence** shows that – when it is in the object position – words like **prove** and **establish** are fairly common, but words like **protest, proclaim, profess, maintain**, and **assert** are even more frequent. (Not to mention the many cases where **innocence** does not appear in this construction at all, e.g. *The issues were not the guilt or innocence of the accused*).[11]

[11] Similar problems arise at **forgiveness**, **approval** (1), **decision**, and **insight**.. COBUILD defines **decision** with an

bundle 5 If someone is **bundled** somewhere, someone pushes them there in a rough and hurried way

We infer that the verb in this use is mainly used in the passive, but corpus data suggests otherwise. Of 877 instances of **bundle**-vb in a 250-million-word corpus there are 101 active uses of the verb in this pattern/meaning, and 109 passive uses. (The count excludes cases where the particles **off** and **up** appear; these are treated separately as phrasal verbs in both COBUILD and MED.) So there is a marked preference for passivization (most transitive verbs passivize less frequently than this), and the entry in MED should in fact have attracted the code [often passive]. But I don't believe this justifies *defining* **bundle** passively and appearing to relegate other uses to the sidelines.

A final example:

cheat If someone **cheats** you out of something, they get it from you by behaving dishonestly (emphasis mine)

This is certainly a common pattern (found in at least 25% of cases where **cheat** takes an object) – but so is 'cheat someone of something' and (more to the point) so is the simple V+O pattern.[12]

The principle that definitions should deal with 'the probable not the possible' is a sound one, and setting the boundary between a frequent pattern and an overwhelmingly marked preference isn't always easy. But the evidence presented here suggests that the requirement of specifying lexical and syntactic environments often leads to defining statements which appear to exclude a wide range of completely regular behaviours. This could be an inherent weakness in the model, and it certainly presents users with a problem of interpretation: for example, when 'prove' is used in the definition of **innocence**, or 'gain' in that of **insight**, should the user infer that this is the only (or the overwhelmingly most frequent) collocate, or simply one of many typical collocates?

4.3 New conventions for old

We have seen that many individual COBUILD definitions may be difficult for learners to process, while others may mislead by overspecifying typical co-texts. But most of the defining *styles* used by COBUILD are transparent (in the sense that you only need to be able to read English in order to understand what they are telling you).[13] There are a few cases, however, of definition styles where the full range of information the lexicographer sets out to convey is only retrievable if the user understands certain conventions unique to COBUILD. I will look briefly at three of these:

---

explanation that begins: 'When you make a **decision**...' – a frequent pattern, to be sure, but representing only around 12%-13% of all uses of this noun. Similarly at **insight**, which starts 'If you gain **insight** or an **insight**...'. But cases where the subject is non-human, and the verbs are **provide, give, offer** etc. are far more common.

[12] The entry at **inform** has similar problems.

[13] Though the same can be said for definitions in other MLDs, which generally now avoid the kind of 'lexicographese' found, say, in 1980s editions of LDOCE and OALD.

• *the If/When distinction*: most verb definitions begin with 'If', but a substantial minority begin with 'When'. For example:

When a horse **gallops**, it runs...

If you **gallop**, you ride a horse that is galloping

The distinction is motivated rather than arbitrary: it is intended to say something to the user (Hanks 1987.126). In most cases I can understand why one is used rather than another (though the entry for **break** has defeated me). I am more or less certain that the average learner (assuming s/he even notices this variation) will not pick up the difference the lexicographer intends.

• *the If you/If someone distinction*: Barnbrook (2002.7-9) compares the entries for **prat** and **bastard**:

If you call someone a **prat**, you mean that they are very stupid or foolish

If someone calls someone else a **bastard**, they are referring to them or addressing them in an insulting way

He notes: 'The difference between the "if you" at the beginning of the definition of "prat" and the "if someone" at the beginning of this definition [**bastard**] is an implicit signal to the user that "bastard" is likely to be regarded as a stronger and more offensive word'. Again, the distinction is well-motivated, but anyone with experience of observing learners using MLDs will probably conclude that this 'implicit signal' is likely to remain implicit.

• *the 'displacement strategy'*: another influential feature of the first COBUILD dictionary was the attention paid to pragmatics and the various ways in which features such as speaker attitude, vagueness, and politeness are typically encoded. Many COBUILD definitions employ what Hanks (1987.133) calls a 'displacement strategy' in order to account for this type of 'meaning'. Compare:

mug 1 A **mug** is a large deep cup with straight sides and a handle, used for hot drinks...

3 If you say that someone is a **mug**, you mean that they are stupid and easily deceived by other people

The 'If you say...' introduction to the second definition exemplifies the displacement strategy, distinguishing this use from the simple denotative meaning signalled by the first definition.[14]

A first observation is that it seems to have been difficult to establish guidelines as to when this style is used. The displacement strategy is employed, for example, at **inflammato-**

[14] See also Sinclair 1987.xvi: 'The words "if you say that..." very often signal metaphoric, figurative, and other non-literal messages'.

ry, controversial, and **confrontational**, but not at **contentious**, **combative,** or **belligerent**; at **scrawny** and **plump**, but not at **svelte** or **chubby**; at **fabulous** but not at **stunning**; and so on. Secondly, this approach (inevitably) exacerbates the problems of length and complexity discussed above. Compare for example these entries for **confrontational**:

if you describe the way that someone behaves as **confrontational**, you are showing your disapproval of the fact that they are aggressive and likely to cause an argument or dispute (COBUILD3)

tending to deal with people in an aggressive way that is likely to cause arguments, rather than discussing things with them (OALD7)

It is not clear that the first definition contains more information; true, it alerts the user to the disapproving attitude, but a learner who encounters this word in context, then decodes it using the second definition, will infer that calling someone 'confrontational' is not a compliment.[15]

It is to COBUILD's credit that they identified a problem which needs addressing – traditional lexicography's failure to account successfully for non-denotative meaning – and started the process of finding solutions. My own view is that these solutions do not deliver: they tend to expand already long definitions, and the information they aim to convey can only be fully understood by a user who has learned how these defining conventions work (since the conventions not transparent).[16] Other dictionaries have tried different approaches (see for example the entries for **just good friends, idle rich**, and **nerd** in MED), and others will judge how successful they are. But this remains an area where more research is needed to find solutions that really work. In approaching this – and any other aspect of defining – there is an important rule to keep in mind: what matters is not the lexicographer's intention but the user's interpretation.

5 More than one way to skin a cat

The FSD model arose in response to a set of objectives (see section 2) which are on the whole very good objectives. The question is whether they lead us, inescapably, to the wholesale abandonment of established lexicographic conventions in favour of the FSD. I would argue, rather, that FSDs represent a useful new strategy to add to other defining styles, but that the same objectives can be often be achieved through other means – and in ways that make fewer demands on learners. Most of the main MLDs make quite frequent use of FSDs, but none except COBUILD uses them all the time. In the Macmillan dictionaries (the same may apply to other MLDs, but their Style Guides are not available to me), we recommend the use of FSDs in the following cases:

---

[15] Meanwhile, the definition for **inflammatory**, which begins 'If you accuse someone of saying or doing **inflammatory** things...', risks overspecifying ('accuse' is a strong word to use here).

[16] The entry for **great** is an interesting case, with several different styles used for explaining the adjectival uses: 'You use **great** in order to', 'You say **great** in order to', 'You can describe someone who...as **great**', 'If you describe someone or something as **great**,' and so on.

5.1 Defining verbs

• verbs (especially intransitive verbs) where it is critical to specify the typical range of *subjects*, for example:

expire if an agreement, offer, or official document expires, the period of time during which it exists or can be used comes to an end (MED)

The same style is used at: **pink** (of car engines), **abdicate, buzz** (if your head is buzzing with ideas...), and many others.

• some reflexive verbs: see for example both definitions of the verb **ally** ('if a country allies itself with another country, ...', 'if you ally yourself with someone...'); and some ergative verbs.

• transitive verbs that occur overwhelmingly in the passive: see for example the entries at **apprentice** ('if someone is apprenticed to another person, ...'), **beach** (of whales), **cheer** ('if you are cheered by something such as a piece of news...'). When passivization is merely frequent (rather than dominant), we tend to define in the active and add the note [often passive].

For most transitive verbs, subject-specification is not especially important, so a conventional style works perfectly well (and is more economical). For example:

assassinate to kill a famous or important person, especially for political reasons or for payment

5.2 Defining adjectives

FSDs tend to work well in the following cases:

• where the range of typical complements is narrow and worth specifying, for example:

blistering blistering criticism is very severe

And similarly at **bounden** (duty), **isotonic** (drinks), **jobbing** (workers), **bouffant** (hair), and many others.

• where the word suggests a permanent characteristic, for example:

argumentative someone who is argumentative often argues or disagrees with people

• where a conventional definition cannot be achieved without unnatural and convoluted wording (adjectives are more problematic in this regard than most other word classes):

slippery a slippery surface, object etc is difficult to move on or hold because it is smooth, wet, or covered in something such as ice or oil

But the majority of adjectives can be handled successfully using conventional styles. Compare for example, these two definitions of **lonesome**:

unhappy because you are alone or because you have no friends (MED)

someone who is **lonesome** is unhappy because they do not have any friends or do not have anyone to talk to (COBUILD3)

A case of 'less is more', perhaps.

5.3 Other strategies

MED occasionally uses FSDs in other situations, when a traditional approach produces a more difficult expression.[17] Lexicographers are allowed to use their discretion but the Style Guide lists a series of conventional defining styles, and these are the recommended default. It is worth adding, finally, that information which COBUILD packs into its FSDs can often be conveyed (arguably more clearly) through a *combination* of a simple definition, a list of frequent collocates, and a set of examples. A case in point is the way the various uses of **badly** are explained in MED:

2 in a serious or severe way: *Her eye was cut quite badly.*| *One of the prisoners had been badly beaten by guards.*| **badly damaged/hurt/injured/wounded**: *Fortunately, none of the drivers was badly hurt.*| **badly hit/affected**: *London is one of the worst-affected areas.*[18]

To conclude on this point: MED – and this probably applies to other MLDs – uses FSDs systematically, for categories that gain most benefit from the approach. In many other cases, however, we prefer an accretive strategy: adding layers of information to a simpler definition, using devices such as add-on sentences, labelling, usage notes, transparent grammar codes, and glossed examples.

6 Conclusions

The iconoclasm of the COBUILD project has been good for lexicography. COBUILD began by junking the whole repertoire of traditional defining practices, and out of this arose an imaginative set of new definition-types which have broadened lexicographers' options. Following the 'thesis-antithesis-synthesis' model, most learners' dictionaries have absorbed these lessons and use FSDs selectively – when they appear to be the most effective strategy. COBUILD also gave fresh impetus to a process already underway in the 1980s – starting perhaps with the *Collins English Dictionary* (1st edition 1979) – in which defining practices were re-evaluated with a view to humanising definitions and bringing them closer to 'normal prose'.[19] Sinclair talks about 'the cryptic messages that are still the most common form of

[17] They are almost never used for defining *nouns*, however: in very many cases, there is nothing useful to say about a noun's contextual preferences, and I am not convinced that the considerable extra length that the FSD approach adds is offset by much genuinely useful information.

[18] A similar approach is used in LDOCE4: see for example its entry for **shoulder** (verb), where collocates like **responsibility, blame** and **burden** are listed ahead of a conventional definition.

[19] For example, the various non-transparent uses of parentheses in definitions of verbs and adjectives – criticised by Hanks (1987.116) – has been abandoned in most MLDs.

definition'(2005.428): this may be a fair description of earlier dictionaries, but it has little relevance to 21st century MLDs. This is a competitive and well-informed market, and if definitions in the other MLDs were all 'cryptic', learners would vote with their wallets and buy only COBUILD dictionaries.

The two papers in which Sinclair discusses FSDs raise important issues that deserve our attention. For example, how long can a definition be before the advantages of fuller information are offset by the difficulty of processing it? As linguistic data becomes more abundant and corpus-querying software more powerful, we can learn more and more about the meanings and typical uses of the words we have to define. This means we face difficult choices about what information to select (knowing what *not* to say is one of the hardest lexicographic skills), and the more help we can get with these decisions, the better. Consider for example this definition for **forge**:

If one person or institution **forges** an alliance or relationship with another, or if two people or institutions **forge** an alliance or relationship, they create it with a lot of hard work, hoping that it will be strong and lasting.

On the one hand, this is a more *informative* definition than anything found elsewhere: it tells us about typical subject-types and typical objects, about the verb's reciprocal use, and about some useful semantic features (hard work, the aim to make something lasting). On the other hand, it is 40 words long, and this is not a trivial objection – some users may baulk at reading it, most will find it difficult to follow. Suppose instead we said:

to create an alliance or relationship through a lot of hard work, hoping that it will be strong and lasting

This is half the length of the COBUILD definition and far easier to read. True, some information is lost, about typical subjects and reciprocity, but does this matter? The former point is, arguably, deducible through users' real-world knowledge ('alliances' are generally made by governments and corporations, not ordinary individuals); the latter could be conveyed using a grammar note; and both can be illustrated in well-chosen example sentences. I am reminded of Hanks' seminal paper on defining (1979), which discusses the notion of 'elegance' in definitions. Though Hanks only hints at what this means in this context, we can think of elegance (by analogy with mathematical proofs or computer programs, for example) as a quality that combines simplicity and economy with effectiveness and adequacy.[20] Some FSDs are elegant in these terms, but too many are not.

Finally, we should consider Sinclair's own hypothesis about the non-adoption of FSDs: 'I have often wondered why full-sentence definitions (FSD) have not appealed more to lexicographers, until I realised recently that most of the additional information they provide comes from studying the corpus evidence' (2005.427f.). In other words, most MLDs do not

[20] Johnson, with characteristic foresight, comes close to defining this concept, when he talks about the need, in definitions, for 'brevity, fulness, and perspicuity' (*Plan* 1747). That combination of brevity and 'fulness' is key.

use FSDs because – unlike COBUILD – they do not really use corpus data (or at least, not properly). This assertion cannot go unchallenged. In Sinclair's view, COBUILD is 'still the only *corpus-driven* dictionary': the contrast is with a *corpus-based* dictionary, following the model proposed by Tognini-Bonelli (2001). As I understand it, a *corpus-based* approach sees the corpus mainly as a repository of examples with which to test and exemplify existing statements about language – rather than (as in a *corpus-driven* methodology) a source of data that can help us to radically reshape inherited descriptions. This sounds like a useful distinction when discussing methodologies for theoretical linguistics.[21] But as a description of what goes on in contemporary MLD lexicography, it is an outdated caricature. Sinclair describes COBUILD's aim as 'to write a dictionary that attempted to represent comprehensively the senses and uses of words and phrases *as they were found in a corpus*' (ibid., author's own emphasis). Would any self-respecting lexicographer want to do anything else?

References

A. Dictionaries

COBUILD 1, 2, 3. *Collins COBUILD English Dictionary*, Editions 1 (1987), 2 (1995), 3 (2001). London & Glasgow, HarperCollins.

LLA. *Longman Language Activator* (1993), Harlow, Longman.

LDOCE 1, 4. *Longman Dictionary of Contemporary English*, Editions 1 (1978), 4 (2003). Harlow, Longman.

MED. *Macmillan English Dictionary for Advanced Learners* (2002), Oxford, Macmillan Education.

OALD 4, 7. *Oxford Advanced Learner's Dictionary*, Editions 4 (1989), 7 (2005), Oxford, OUP.

B. Other literature

Barnbrook, G. (2002), *Defining Language: A local grammar of definition sentences*, Amsterdam, John Benjamins.

Hanks, P. (1979), 'To what Extent does a Dictionary Definition Define?', in Hartmann, R. R. K. (ed.), *Dictionaries and their Users*. Exeter: University of Exeter, pp. 32-38.

Hanks, P. (1987), 'Definitions and Explanations' in Sinclair, J. M. (ed.), *Looking Up*. London, Collins.

Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). 'The Sketch Engine', in Williams, G. and Vessier, S. (eds), *Proceedings of the Eleventh EURALEX Congress*. Lorient, France, Université de Bretagne-Sud, pp. 105-116.

Rundell, M. (1998), 'Recent trends in English pedagogical lexicography', International Journal of Lexicography 11.4, pp. 315-342.

Sinclair, J. (1987), Editor's Introduction to COBUILD1.

Sinclair, J. (2004), 'In praise of the dictionary', in Williams, G. and Vessier, S. (eds), *Proceedings of the Eleventh EURALEX Congress*. Lorient, France, Université de Bretagne-Sud, pp. 1-12.

Sinclair, J. (2005), 'To complement the dictionary', in Blatná, R. and Petkevič, V. (eds) *Jazyky a jazykov da: Sborník k 65. narozeniná prof. Františka Čermá*. Prague, Charles University, pp. 417-444.

Stock, P. (1988), 'The structure and function of definitions', in Snell-Hornby, M. (ed.), *ZuriLEX '86 Proceedings*. Tübingen, Franke Verlag, pp. 81-90.

[21] And it is actually a fair description of a transitional period (late 1980s-early 1990s) when established MLDs (OALD and LDOCE) grappled with the challenge of meshing the new insights from corpora with dictionaries that had been wholly written during in the pre-corpus era.